GlySim: Modeling and Simulating Glycemic Response for Behavioral Lifestyle Interventions

Asiful Arefeen, and Hassan Ghasemzadeh

***Abstract*—Effective prevention and management of diabetes relies on maintaining a normal blood glucose level, thus avoiding abnormal events such as hyperglycemia and hypoglycemia. Predicting anomalous events beforehand can potentially help patients and caregivers intervene to prevent such events through modifiable behaviors such as exercise, diet, and medication. Although Continuous Glucose Monitor (CGM) sensors have been used to monitor and forecast blood glucose level, current research lacks a computational approach that recommends a behavioral intervention to bring the glucose level to a normal range. To address this shortcoming, we present *GlySim*[1], a CGM simulator that uses multimodal data to not only forecast future glucose readings but also enable a user to examine the impacts of behavior change on glucose response in advance. GlySim creates opportunities for change in food consumption, medication, and physical activity to avoid dysglycemia by pinpointing factors that cause anomalous events using Grad-CAM (Gradient-weighted Class Activation Mapping) and allowing users to observe how adjusting a behavioral factor changes glucose trajectories. We validate GlySim on a dataset of 10 patients with type 1 diabetes and achieve an overall mean absolute error (MAE) as low as 16.5 mg/dl in simulating glycemic response. Furthermore, GlySim detects hyperglycemic events with 0.89 average precision.**

***Index Terms*—Wearable, continuous glucose monitor, diabetes, forecasting, multimodal data, simulator, digital twin**

I. Introduction

Glucose control, as well as diet monitoring and management [1], is critical in both diabetes prevention and diabetes management. Poor glucose control leads to abnormal events such as hyperglycemia and hypoglycemia. Regular exposure to dysglycemia can increase the risk of complications, including cardiovascular disease, reduced eyesight, cancer, impaired glycemic control, and seizure [2]. As part of remote health monitoring and human-centered IoT applications [3], [4], CGM sensors are commonly used to measure blood glucose. Although there exist technologies for diet [5] and glucose monitoring, their utility for behavior change is currently limited. Current CGM systems lack computational capabilities to actively influence users' behavior and work as a decision support for individuals with or at risk for diabetes. Such a limitation thwarts the effectiveness of CGM systems to proactively promote health enhancing behaviors associated with food consumption, medication, and physical activity, to maintain stable blood glucose levels.

Recent work developed algorithms to predict blood glucose from CGM sensors [6]. One study used a moving window on blood glucose and accelerometer signals to predict future blood glucose levels with a LSTM (long short-term memory) network [7]. Another study developed a lightweight LSTM model, aiming to deploy it on a microcontroller for enhanced practicality [8]. Additionally, a low-cost and low-power wearable system was designed with attention-based evidential recurrent neural network convenient for use in clinical settings [9]. Another study used multimodal data in a stacked LSTM model followed by Kalman filter to predict blood glucose values [10]. Furthermore, a modified ResNet architecture was introduced to predict blood glucose levels 30 minutes into the future [11]. Although the value of accurately forecasting blood glucose level can be perceived from these studies, little effort is given in designing a mechanism that enables interventions to prevent abnormal blood glucose levels such as hyperglycemia and hypoglycemia.

We address these shortcomings by developing *GlySim*, a simulator for glycemic response that takes data from multimodal information sources and allows users to virtually examine the effects of changes in behavioral parameters on blood glucose. Developing methods for modeling and simulating glycemic response is central to the development of a digital twin for human health. Beyond effective diabetes prevention and management, a glycemic simulator will provide a platform for developing and testing novel algorithms and techniques for glucose management without extensive clinical studies. Such a simulator will also be instrumental for developing a reinforcement learning aided insulin optimizer, where the reward function is based on maintaining a normal blood glucose level. Simulating glucose response, researchers and developers can test the effectiveness of different interventions, algorithms, and decision-making strategies in a controlled and flexible virtual environment. Moreover, GlySim enables the construction of large-scale and diverse datasets to educate and train healthcare professionals, patients, caregivers and AI models on the interpretation of CGM data in a risk-free environment.

GlySim can be viewed as a deep learning based blood glucose simulator that enables us to predictively conduct virtual interventions. In designing GlySim, a neural network is first introduced to predict future glucose readings over a prediction horizon and to identify occurrences of impending dyslycemia. A dashboard simulator is then designed to visualize the predicted signals and allow the user to tune behavioral parameters and observe the impact on the glucose curve. An intervention strategy is finally recommended by using gradients of each target prediction to determine what behavioral factor is most

Authors are with the College of Health Solutions, Arizona State University, Phoenix, AZ 85004, USA email: aarefeen@asu.edu

[1]Resources available at: https://github.com/Arefeen06088/GlySim

responsible for the predicted outcome, followed by an iterative process to virtually determine the intervention impact.

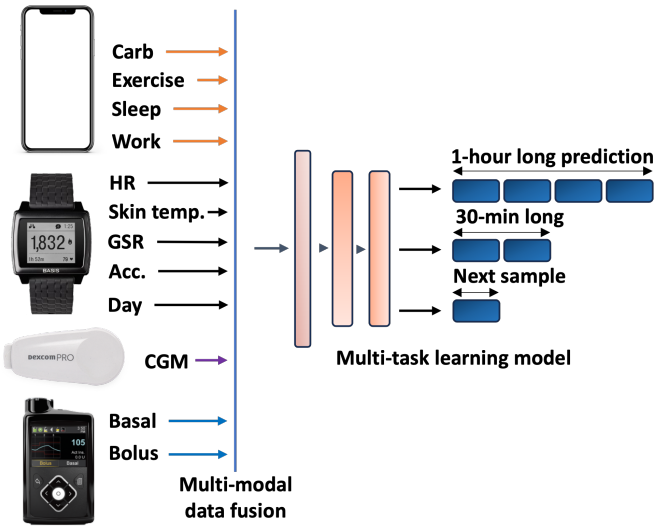

Fig. 1: GlySim integrates data from smartphone, smartwatch, wearable CGM and insulin pump, and devises a multi-task learning model to predict glucose response.

II. GlySim Design

GlySim consists of two main modules, prediction and intervention, which respond to the modeling and simulation requirement of a digital twin, respectively.

Fig. 1 shows a high level diagram that illustrates the prediction approach. To this end, we formulate a predictive model that utilizes a neural architecture to forecast future blood glucose levels based on multimodal data. Let $X = \{x^{(1)}, x^{(2)}, ..., x^{(d)}\}$ be the set of d feature/sensor observations where each observation $x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, ..., x_t^{(1)}\}$ is of length t. We aim to predict the future blood glucose levels within a specific interval, referred to as prediction horizon. If n represents the total number of blood glucose readings to be predicted, our target variable will be $Y = \{y_1, y_2, ..., y_n\}$. Hence, our goal is to learn a predictive model that can estimate the future blood glucose levels given the input features. Mathematically, we seek a function f that maps the input features X to the predicted blood glucose levels Y:

$$f : \mathbb{R}^{d \times t} \rightarrow \mathbb{R}^n \tag{1}$$

where $\mathbb{R}^{d \times t}$ represents the d-dimensional feature space of length t and $\mathbb{R}^n$ represents the n-dimensional space of the output, for the predicted blood glucose levels.

In addition to the prediction module, GlySim has an intervention module that allows one to simulate virtual interventions. If the predicted blood glucose level is above a hyperglycemic threshold, i.e. $\max(f(X)) \geq 178$, an intervention can be recommended to prevent the impending hyperglycemia.

A. Predictive Modeling Using Deep Learning

For glucose level prediction, we devise a 3-layer stacked CNN-LSTM network architecture in GlySim (Fig. 2). The

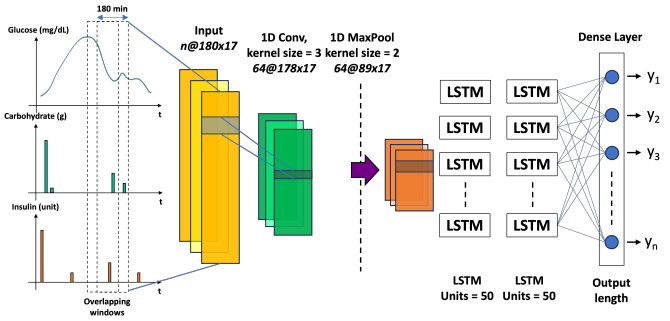

Fig. 2: Stacked CNN-LSTM architecture of GlySim. Input to the model is 3-hour long multimodal data.

ideal approach to realize such as architecture is to apply a sliding window on the multivariate time series and slice it to produce sequences of length t for input batches X. The convolutional layer, with $tanh$ activation, applies filters to the input data to extract high-level relevant features. Therefore, if $\mathbf{W}_{conv}$ and $\mathbf{b}_{conv}$ refer to the weights and biases of the convolution layer, the output can be written as-

$$\mathbf{h}_{conv} = \tanh(\mathbf{W}_{conv} * X + \mathbf{b}_{conv}) \tag{2}$$

The LSTM layer processes the sequential data, capturing both short-term and long-term dependencies. It takes the sequence of feature maps produced by the convolutional layer as input. The LSTM layer maintains a memory cell that can store and update information over time. The layer uses gates (input, forget, and output) to control the flow of information within the memory cell. The LSTM layer can learn to remember or forget specific information based on the patterns in the input sequence. By stacking multiple LSTM layers, the network can capture more complex temporal dependencies. If o_t is from final output gate and c_t is from final cell state, the output of LSTM layer can be written as-

$$\mathbf{h}_{LSTM} = o_t * \tanh(c_t) \tag{3}$$

The output layers of the model are configured to generate multi-task outputs, where each task corresponds to a different prediction horizon (i.e., 5-minute, 30-minute, 60-minute).

B. Intervention Simulation

Beside forecasting future blood glucose levels with the predictive model, we introduce an approach to simulate interventions and identify a way that can lead to preventing dysglycemia (e.g., a hyperglycemic condition). Once the model predicts values above the hyperglycemic threshold, an iterative process is triggered to determine the required reduction in consumed CHO amount from the input data in order to maintain blood glucose levels within the desired range.

We manipulate the consumed $carbohydrate$ (or any other parameter effecting blood glucose) amount in input feature X in an iterative process to identify the optimal reduction in CHO to achieve a predicted CGM value below hyperglycemic threshold. Algorithm 1 is for further illustration.

With this intervention mechanism in place, the aim is to optimize the consumed CHO amount and maintain blood

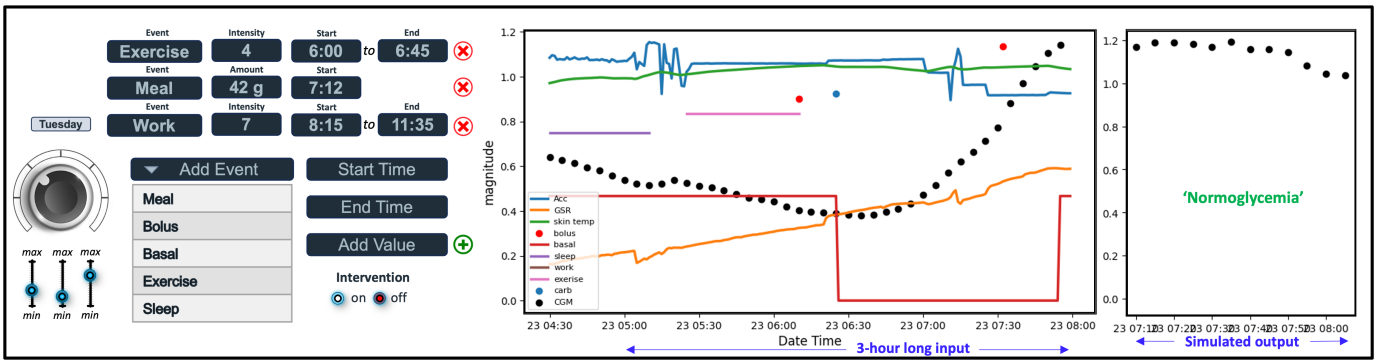

Fig. 3: GlySim dashboard for simulating glucose response based on behavioral parameters. The dashboard will hold options to add events like meal, exercise, work, sleep etc. Values are scaled for visualization.

Algorithm 1 CHO reduction based intervention

Inputs:
multimodal features, $X' = \{x^{(1)}, x^{(2)}, ..., x^{(d-1)}\}$
projected CHO intake, C
hyperglycemic threshold, $BG_{\max}$
predictive model, $f(\cdot)$
Output: CHO reduction, C_r
if $\max(f(X \oplus C)) \geq BG_{\max}$ **then**
  while $\max(f(X \oplus C - C_r)) \geq BG_{\max}$ **do**
    $C_r += 1$
  end while
return C_r
else
  $C_r = 0$
end if

glucose levels within the desired range for effective glucose management.

III. Experimental Validation

A. Dataset

We chose the OhioT1DM [12] to demonstrate the results of our proposed methods. OhioT1DM was collected over an eight-week long clinical study of 12 deidentified T1 diabetes patients. Participants wore Medtronic 530G/630G insulin pumps and Medtronic Enlite CGM that transmits blood glucose level every 5 minutes. Their physiological data (acceleration, skin response etc.) was recorded on either Basis Peak fitness band or Empatica Embrace while their self-reported CHO intake, work, exercise intensity and sleep quality were recorded on smartphones. Of these eight weeks, roughly 44 and 12 days were allocated for train and test respectively. Missing values were imputed with interpolation or extrapolation. However, subjects 540 and 567 were excluded from analysis due to not having entries for CHO, work, sleep, and exercise sessions.

B. Model Development

After smoothing the CGM stream with Kalman filter, we produced 180-minute long multimodal time-series windows with 175-minute overlap and fed them to a network with one 1-D CNN layer and two LSTM layers, all equipped with $tanh$ activation and followed by a 50% droptout. The CNN layer had 64 filters, a kernel size of 3, l_2 regularizer followed by a $maxpool$. The two LSTM layers had 50 units each. The output specific layers were for 60-minute, 30-minute and immediate next sample prediction each with $linear$ activation. Adam optimizer was used with a learning rate of 0.01 and a decay of 0.001. We also used Grad-CAM [13] in the intervention phase to identify the responsible factor(s) behind hyperglycemic/hypoglycemic events.

C. Results

The performance of the simulator was measured using mean absolute error (MAE) and root mean squared error (RMSE) shown in Equations (4) and (5).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

where, y_i, $\hat{y}_i$ and n refer to actual value, predicted value and number of test samples respectively.

Table I summarizes the performance of the proposed approach across different subjects and prediction horizons. Regression performance varies across different prediction horizons. While the model recorded the lowest MAE and RMSE

TABLE I: MAE (mg/dL) and RMSE (mg/dL) of blood glucose level estimation for different subjects and prediction horizons.

Subject	60-minute		30-minute		next sample	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
544	15.7	22.6	12	16.5	9	11.5
552	9.9	17.9	7.7	13.1	5.6	8.9
559	18.6	27	14.2	20.1	10.3	14.7
563	18.8	26.3	14.4	20.1	9.8	13.7
570	17.3	24.2	13.4	18.5	9.5	13.4
575	18	26.3	13.2	19.2	8.9	12
584	18.6	27.5	13.4	19	7.8	10.3
588	16.8	24.5	12.4	18	8.5	12
591	18	25.2	12.5	17.5	6.3	8.6
596	15.4	22.8	11	16	6.7	9.5
average	16.5	24.2	12.3	17.5	8.2	11.3
Agnostic	20.5	27.3	17.6	21.8	15.3	18.4

TABLE II: Performance of the models in detecting hyperglycemic events.

Subject →	**544**	**552**	**559**	**563**	**570**	**575**	**584**	**588**	**591**	**596**
precision	0.903	0.8	0.913	1	0.885	0.969	1	0.714	0.875	0.871

values for the 60-minute and 30-minute prediction horizons for subject 591, the lowest next sample prediction error was found for subject 584. The average MAE and RMSE values serve as a summary measure of the overall estimation accuracy. In addition to subject specific models, we also trained a subject agnostic model which, as expected, incurred higher error.

Performance of the models to detect hyperglycemia is summarized in Table II using precision.

$$\text{precision} = \frac{TP}{TP + FP}$$

where, TP and FP refer to true positives and false positives. The prediction model can detect hyperglycemia with an average precision of 0.89 (min 0.71, max 1). The intervention simulator kicks in if any of the predicted values exceeds hyperglycemic threshold. Initially, the algorithm identifies the factor causing hyperglycemia or hypoglycemia. It determines whether the abnormal blood sugar level is due to the consumed carbohydrate (CHO) amount or if it is caused by factors like insulin intake or exercise leading to hypoglycemia. The Grad-CAM technique helps to pinpoint the underlying factor contributing to the rise or fall in blood glucose levels. Once the key factor is identified, the proposed algorithm iteratively adjusts its value to minimize the impact and prevent the anomalous event from taking place.

Fig. 4 provides a demonstration of the designed intervention for subject 588. As soon as the model predicts values exceeding the hyperglycemic threshold, the intervention initiates by employing Grad-CAM to locate the factor responsible.

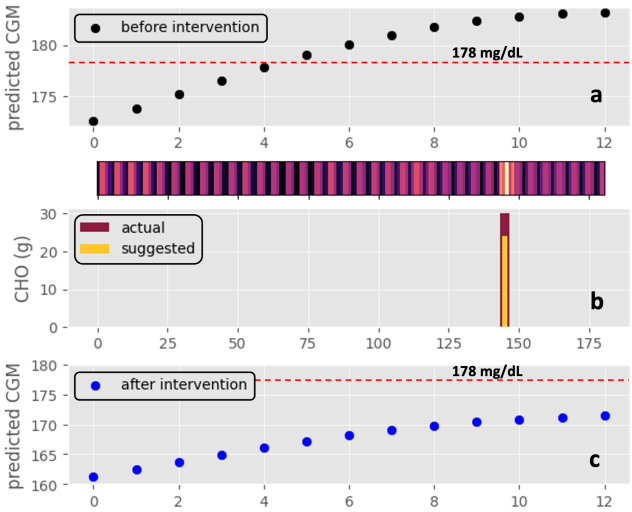

Fig. 4: Intervention phase identifies the key factor for hyperglycemia (or hypoglycemia) using Grad-CAM. Then repeatedly changes the value to identify the optimal amount that ensures normoglycemia.

As shown, the heatmap produced by Grad-CAM highlights that the consumption of $30g$ of CHO might be causing hyperglycemia. Now, the algorithm iteratively reduces the CHO amount and determines a reduced value of $23g$ ensures predictions below the threshold. Table III displays the iterations, reduced CHO amounts and corresponding maximum predicted blood glucose levels.

TABLE III: Results from iterative intervention process.

Iteration #	Reduced CHO (g)	Resulting maximum predicted CGM
1	30	183.2
2	29	182.8
3	28	182.4
4	27	181.6
5	26	180.2
6	25	179.5
7	24	178.7
8	**23**	177.5
9	22	176.9

In other cases/examples, Grad-CAM also detects whether previous high glucose levels contribute to hyperglycemia or if exercise intensity or insulin lead to hypoglycemia. However, when hyperglycemia persists for over 3 hours, the simulator naturally cannot recommend an intervention to prevent the event. Additionally, due to the limitations of the dataset, there are instances where subjects experience hyperglycemia despite not consuming any CHO or being on insulin. In such cases, the algorithm struggles in generating suggestions as there are no concrete modifiable parameters to regulate their blood glucose.

IV. Limitations and Future Works

GlySim presents one approach to simulating behavioral factors and observing the impact of behavior changes on glucose response. A potential application area of such a tool is in designing a digital twin library to simulate human blood glucose level in presence of different constraints like specific age, gender, meal and insulin intake and to use the technology in clinical decision making and in identification of optimal behavioral treatments. However, our progress is currently in early stages. Our objective is to create a simulator featuring a dashboard similar to the one depicted in Fig. 3. Achieving such a goal requires effort from both a machine learning standpoint to enhance model robustness against corner cases and from a software development perspective to design a user-friendly and interactive interface. Furthermore, deploying an AI-driven intervention in a clinical environment imposes greater challenges developing such as technology. The entire process may include phases like approval, validation of scalability, reliability, user-adaptability, regulatory compliance, and maintenance.

As indicated in Table I, the subject-agnostic model is not as effective as the subject-specific models, leading to

generalization problems of the predictive model aided by inconsistent settings and diverse subjects [14]. Therefore, part of our follow-up work includes applying transfer learning and domain adaptation techniques [15] to improve the generalization performance of the subject-independent model.

V. Conclusion

We introduce GlySim, a framework for modeling blood glucose response and simulating behavioral interventions based on multimodal behavioral, physiological, and health data. GlySim provides an extended forecast horizon utilizing the multimodal data. The intervention simulation mechanism is designed to empower individuals to make behavior changes related to food consumption, medication intake, and physical exercise. The experimental validation on ten individuals demonstrates promising performance in terms of prediction accuracy, with low MAE and RMSE values for different prediction horizons, as well as potential for identifying the optimal lifestyle intervention. One of our future endeavors is to use this simulator to develop a reinforcement learning based intervention system and deploy the system in clinical studies.

Acknowledgment

This work was supported in part by the National Science Foundation, under grant CNS-2210133. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organization.

References

[1] A. Arefeen, N. Jaribi, B. J. Mortazavi, and H. Ghasemzadeh, "Computational framework for sequential diet recommendation: Integrating linear optimization and clinical domain knowledge," *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 91–98, 2022.

[2] A. Arefeen, S. N. Fessler, C. Johnston, and H. Ghasemzadeh, "Forewarning postprandial hyperglycemia with interpretations using machine learning," *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4, 2022.

[3] S. A. Rokni and H. Ghasemzadeh, "Plug-n-learn: Automatic learning of computational algorithms in human-centered internet-of-things applications," *2016 53nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, pp. 1–6, 2016.

[4] S. I. Lee, H. Ghasemzadeh, B. Mortazavi, M. Lan, N. Alshurafa, M. Ong, and M. Sarrafzadeh, "Remote patient monitoring: what impact can data analytics have on cost?" in *Proceedings of the 4th Conference on Wireless Health*, 2013, pp. 1–8.

[5] N. Hezarjaribi, S. Mazrouee, and H. Ghasemzadeh, "Speech2health: A mobile framework for monitoring dietary composition from spoken data," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 252–264, 2018.

[6] H. Nemat, H. Khadem, M. R. Eissa, J. Elliott, and M. Benaissa, "Blood glucose level prediction: Advanced deep-ensemble learning approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, pp. 2758–2769, 2022.

[7] W. P. T. M. van Doorn, Y. D. Foreman, N. C. Schaper, H. H. Savelberg, A. Koster, C. J. H. van der Kallen, A. Wesselius, M. T. Schram, R. M. A. Henry, P. C. Dagnelie, B. E. de Galan, O. Bekers, C. D. A. Stehouwer, S. J. R. Meex, and M. C. Brouwers, "Machine learning-based glucose prediction with use of continuous glucose and physical activity monitoring data: The maastricht study," *PLoS ONE*, vol. 16, 2021.

[8] T. Zhu, L. Kuang, K. Li, J. Zeng, P. Herrero, and P. Georgiou, "Blood glucose prediction in type 1 diabetes using deep learning on the edge," *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2021.

[9] T. Zhu, L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou, "Iomt-enabled real-time blood glucose prediction with deep learning and edge computing," *IEEE Internet of Things Journal*, vol. 10, pp. 3706–3719, 2023.

[10] M. F. Rabby, Y. Tu, M. I. Hossen, I. Lee, A. Maida, and X. S. Hei, "Stacked lstm based deep recurrent neural network with kalman smoothing for blood glucose prediction," *BMC Medical Informatics and Decision Making*, vol. 21, 2021.

[11] H. Rubin-Falcone, I. Fox, and J. Wiens, "Deep residual time-series forecasting: Application to blood glucose prediction," in *KDH@ECAI*, 2020.

[12] C. R. Marling and R. C. Bunescu, "The ohiot1dm dataset for blood glucose level prediction: Update 2020," *CEUR workshop proceedings*, vol. 2675, pp. 71–74, 2020.

[13] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, pp. 336–359, 2016.

[14] P. Alinia, A. Arefeen, Z. E. Ashari, S. I. Mirzadeh, and H. Ghasemzadeh, "Model-agnostic structural transfer learning for cross-domain autonomous activity recognition," *Sensors (Basel, Switzerland)*, vol. 23, 2023.

[15] Y. Ma, A. T. Campbell, D. J. Cook, J. C. Lach, S. N. Patel, T. Ploetz, M. Sarrafzadeh, D. Spruijt-Metz, and H. Ghasemzadeh, "Transfer learning for activity recognition in mobile health," *ArXiv*, vol. abs/2007.06062, 2020.